REMARKS

The application is believed to be in condition for allowance because the claims are novel and non-obvious over the cited art. The following paragraphs provide the justification for these beliefs. In view of the following reasoning for allowance, the applicant hereby respectfully requests further examination and reconsideration of the subject application.

The Rejection of Claims 1-3, 5-6, 14, 18-19 and 23-24 Under 35 USC 102(b).

Claims 1-3, 5-6, 14, 18-19 and 23-34 stand rejected under 35 USC 102(b) as being anticipated by Foote et al. U.S. Patent No. 6,404,925 (hereinafter Foote). It was contended in the above-identified Office Action that Foote teaches all the elements of the rejected claims. The applicants respectfully traverse this contention of anticipation.

The applicants claim a technique that can extract objects from an image sequence using the constraints on their motion and also performs tracking while the appearance models are learned. **The technique operates in near real time, processing data and learning generative models at substantially the same rate/time the input data is received. (Summary)**

The claimed technique tries to recognize patterns in time (e.g., finding possibly recurring scenes or objects in an image sequence), and in order to do so attempts to model the process that could have generated the pattern. It uses the possible states or classes, the probability of each of the classes being in each of the states at a given time and a state transition matrix that gives the probability of a given state given that state at a previous time. The states further may include observable states and hidden states. In such cases the observed sequence of states is probabilistically related to the hidden process. The processes are modeled using a transformed Hidden Markov model (THHM) where there is an underlying hidden Markov process changing over time, and a set of observable states which are related somehow to the hidden states. The connections between the hidden states and the observable states represent the probability of generating a particular

observed state given that the Markov process is in a particular hidden state. All probabilities entering an observable state will sum to 1. (Summary)

The number of classes of objects and an image sequence is all that must be provided in order to extract objects from an image sequence and learn their generative model (e.g., a model of how the observed data could have been generated). Given this information, probabilistic inference and learning are used to compute a single set of model parameters that represent either the video sequence processed to that point or the entire video sequence. These model parameters include the mean appearance and variance of each class. The probability of each class is also determined. (Summary)

More specifically, the applicants claim,

> "A system for automatically decomposing an image sequence, comprising a computer-readable storage medium storing a program that when executed causes:
> a computer to perform the following process actions,
> providing an image sequence of at least one image frame of a scene;
> providing only a preferred number of classes of objects to be identified within the image sequence;
> automatically decomposing the image sequence into the preferred number of classes of objects, using probabilistic inference and learning to compute a single set of model parameters comprising a mean visual appearance and variance of each class in the image sequence, processing the provided image sequence and computing the single set of model parameters at a substantially same time that the image sequence is provided, wherein automatically decomposing the image sequence into the preferred number of object classes comprises performing a probabilistic variational expectation-maximization analysis, comprising:
> forming a probabilistic model having variational parameters representing posterior distributions;
> initializing said probabilistic model;
> inputting an image frame from the image sequence;
> computing a posterior given observed data in said image sequence; and
> using the posterior of the observed data to update the probabilistic model parameters."

And,

> "A computer-implemented process for automatically generating a representation of an object in at least one image sequence, comprising a computer-readable storage medium storing a program:
> that when executed causes a computer to,

acquire at least one image sequence, each image sequence having at least one image frame;

automatically decompose each image sequence into a generative model with each generative model comprising a set of model parameters comprising a mean visual appearance and variance of each class in the image sequence being decomposed, using an expectation-maximization analysis that employs a Viterbi analysis, wherein each generative model is computed at a <u>substantially same time</u> that the at least one image sequence is acquired, <u>wherein an expectation step of the generalized expectation-maximization analysis maximizes a lower bound on a log-likelihood of each image frame by inferring approximations of variational parameters</u>."

Foote discloses methods for segmenting audio-video recording of meetings containing slide presentations by one or more speakers. These segments serve as indexes into the recorded meeting. If an agenda is provided for the meeting, these segments can be labeled using information from the agenda. The system automatically detects intervals of video that correspond to presentation slides. **Under the assumption that only one person is speaking during an interval when slides are displayed in the video, possible speaker intervals are extracted from the audio soundtrack by finding these regions.** Since the same speaker may talk across multiple slide intervals, **the acoustic data from these intervals is clustered to yield an estimate of the number of distinct speakers and their order. Clustering the audio data from these intervals yields an estimate of the number of different speakers and their order. Merged clustered audio intervals corresponding to a single speaker are then used as training data for a speaker segmentation system.** Using speaker identification techniques, <u>**the full video is then segmented into individual presentations based on the extent of each presenter's speech**</u>. (Abstract)

As for Claim 1, and its dependents, Foote does not teach the applicants' claimed <u>automatically decomposing an image sequence into the preferred number of object classes by performing a probabilistic variational expectation-maximization analysis that operates by: forming a probabilistic model having variational parameters representing posterior distributions; initializing the probabilistic model; inputting an image frame from the image sequence; computing a posterior given observed data in said image sequence; and using the posterior of the observed data</u>

to update the probabilistic model parameters. Nor does Foote teach the applicant's claimed number of classes of objects to be identified within the image sequence or automatically decomposing the image sequence into the preferred number of classes of objects, processing data and learning generative models at substantially the same time that the input data is received.

As for Claim 23, and its dependents, Foote does not teach the applicant's claimed automatically decomposing each image sequence into a generative model with each generative model having a set of model parameters that include a mean visual appearance and variance of each class in the image sequence being decomposed. The composition of the image sequence employs an expectation-maximization analysis that includes a Viterbi analysis. Each generative model is computed at substantially the same time that the image sequence is acquired, and an expectation step of the generalized expectation-maximization analysis maximizes a lower bound on a log-likelihood of each image frame by inferring approximations of variational parameters.

Thus, the applicants have claimed an element not taught in Foote. As such, the rejected claims, as amended, are not anticipated by the reference. It is, therefore, respectfully requested that the rejection of Claims 1-3, 5-6, 14, 18-19 and 23-34 be reconsidered based on the above-quoted distinguishing claim language.

The 35 USC 103(a) Rejection of Claims 4, 7 and 27.

Claims 4, 7 and 27 were rejected under 35 USC 103(a) as unpatentable over Foote, in view of Petrovic et al (Transformed Hidden Markov Models: Estimating Mixture Models of Images and Inferring Spatial Transformations in Video Sequences, Computer Visions and Pattern Recognition, 2000, Vol. 2, pg 16-33), hereinafter Petrovic. The Office Action contended that Foote teaches all of the limitations of Claims 4, 7 and 27, except that Foote does not teach a model that employs a latent image and a translation variable in learning each object class, nor does Foote teach using a latent image and a translation variable in filling in hidden variables. However, the Office Action contended that Petrovic teaches these features, rendering Claims 4, 7 and 27 obvious. The applicants respectfully traverse this contention of obviousness.

In order to deem the applicant's claimed invention unpatentable under 35 USC 103, a prima facie showing of obviousness must be made. To make a prima facie showing of obviousness, all of the claimed elements of an applicant's invention must be considered, especially when they are missing from the prior art. If a claimed element is not taught in the prior art and has advantages not appreciated by the prior art, then no prima facie case of obviousness exists. The Federal Circuit court has stated that it was error not to distinguish claims over a combination of prior art references where a material limitation in the claimed system and its purpose was not taught therein (*In Re Fine*, 837 F.2d 107, 5 USPQ2d 1596 (Fed. Cir. 1988)).

As discussed above, the applicants claim,

"A system for automatically decomposing an image sequence, comprising a computer-readable storage medium storing a program that when executed causes:

a computer to perform the following process actions,

providing an image sequence of at least one image frame of a scene;

providing only a preferred number of classes of objects to be identified within the image sequence;

automatically decomposing the image sequence into the preferred number of classes of objects, using probabilistic inference and learning to compute a single set of model parameters comprising a mean visual appearance and variance of each class in the image sequence, processing the provided image sequence and computing the single set of model parameters at a substantially same time that the image sequence is provided, wherein automatically decomposing the image sequence into the preferred number of object classes comprises performing a probabilistic variational expectation-maximization analysis, comprising:

forming a probabilistic model having variational parameters representing posterior distributions;

initializing said probabilistic model;

inputting an image frame from the image sequence;

computing a posterior given observed data in said image sequence; and

using the posterior of the observed data to update the probabilistic model parameters."

And,

"A computer-implemented process for automatically generating a representation of an object in at least one image sequence, comprising a computer-readable storage medium storing a program:

that when executed causes a computer to,

acquire at least one image sequence, each image sequence having at least one image frame;

automatically decompose each image sequence into a generative model with each generative model comprising a set of model parameters comprising a mean visual appearance and variance of each class in the image sequence being decomposed, using an expectation-maximization analysis that employs a Viterbi analysis, wherein each generative model is computed at a substantially same time that the at least one image sequence is acquired, wherein an expectation step of the generalized expectation-maximization analysis maximizes a lower bound on a log-likelihood of each image frame by inferring approximations of variational parameters."

As discussed above, as for Claim 1 and dependents 4 and 7, Foote does not teach the applicants' claimed automatically decomposing the image sequence into the preferred number of object classes by performing a probabilistic variational expectation-maximization analysis, wherein the variational expectation maximization analysis comprises: forming a probabilistic model having variational parameters representing posterior distributions; initializing the probabilistic model; inputting an image frame from the image sequence; computing a posterior given observed data in the image sequence; and using the posterior of the observed data to update the probabilistic model parameters. Nor does Foote teach the applicant's claimed number of classes of objects to be identified within the image sequence or automatically decomposing the image sequence into the preferred number of classes of objects, processing data and learning generative models at substantially the same time that the input data is received. Petrovic also does not teach these features.

As for Claim 23, and its dependent claim 27, Foote does not teach the applicant's claimed automatically decomposing each image sequence into a generative model with each generative model comprising a set of model parameters that have a mean visual appearance and variance of each class in the image sequence being decomposed, by using an expectation-maximization analysis that employs a Viterbi analysis where each generative model is computed at substantially the same time that the image sequence is acquired, and an expectation step of the generalized expectation-maximization analysis maximizes a lower bound on a log-likelihood of each image frame by inferring approximations of variational parameters. Petrovic also does not teach these features.

Accordingly, Foote in combination with Petrovic does not teach the applicant's claim limitations. Nor does Foote in combination with Petrovic recognize the advantages of the applicants' claimed invention. Namely, Foote in combination with Petrovic does not teach allowing video sequences to be decomposed into a preferred number of classes in real-time with a minimal amount of input data. Thus, the applicants have claimed elements not taught in the cited art and which have advantages not recognized therein. Accordingly, no prima facie case of obviousness has been established in accordance with the holding of *In Re Fine*. This lack of prima facie showing of obviousness means that the rejected claims are patentable under 35 USC 103 over Foote in view of Petrovic. As such, it is respectfully requested that Claims 4, 7 and 27 be allowed based on the previously-quoted claim language.

The 35 USC 103(a) Rejection of Claims 20-21 and 25-26.

Claims 20-21 and 25-26 were rejected under 35 USC 103(a) as unpatentable over Foote, in view of Jojic et al (Learning Flexible Sprites in Video Layers, Proc. Of IEEE Conf. on Computer Vision and Pattern Recognition, 2001, pg. 1-8). The Office Action contended that Foote teaches all of the limitations of claims, except that Foote does not various model parameters of the applicants' claimed invention. However, the Office Action contended that Jojic teaches these features, rendering Claims 20-21 and 25-26 obvious. The applicants respectfully disagree with this contention of obviousness.

As discussed above, the applicants claim,

"A system for automatically decomposing an image sequence, comprising a computer-readable storage medium storing a program that when executed causes:

a computer to perform the following process actions,

providing an image sequence of at least one image frame of a scene;

providing only a preferred number of classes of objects to be identified within the image sequence;

automatically decomposing the image sequence into the preferred number of classes of objects, using probabilistic inference and learning to compute a single set of model parameters comprising a mean visual appearance and variance of each class in the image sequence, processing the provided image sequence and computing the single set of model parameters at a substantially same time that the image sequence is provided, wherein

automatically decomposing the image sequence into the preferred number of object classes comprises performing a probabilistic variational expectation-maximization analysis, comprising:

forming a probabilistic model having variational parameters representing posterior distributions;

initializing said probabilistic model;

inputting an image frame from the image sequence;

computing a posterior given observed data in said image sequence; and

using the posterior of the observed data to update the probabilistic model parameters."

And,

"A computer-implemented process for automatically generating a representation of an object in at least one image sequence, comprising a computer-readable storage medium storing a program:

that when executed causes a computer to,

acquire at least one image sequence, each image sequence having at least one image frame;

automatically decompose each image sequence into a generative model with each generative model comprising a set of model parameters comprising a mean visual appearance and variance of each class in the image sequence being decomposed, using an expectation-maximization analysis that employs a Viterbi analysis, wherein each generative model is computed at a substantially same time that the at least one image sequence is acquired, wherein an expectation step of the generalized expectation-maximization analysis maximizes a lower bound on a log-likelihood of each image frame by inferring approximations of variational parameters."

As discussed above, as for Claim 1 and dependents 20-21, Foote does not teach the applicants' claimed automatically decomposing the image sequence into the preferred number of object classes by performing a probabilistic variational expectation-maximization analysis, wherein the variational expectation maximization analysis comprises: forming a probabilistic model having variational parameters representing posterior distributions; initializing the probabilistic model; inputting an image frame from the image sequence; computing a posterior given observed data in the image sequence; and using the posterior of the observed data to update the probabilistic model parameters. Nor does Foote teach the applicant's claimed number of classes of objects to be identified within the image sequence or automatically decomposing the image sequence into the preferred number of classes of objects, processing data and learning generative models at substantially the same time that the input data is received. Jojic also does not teach these features.

As for Claim 23, and its dependent claims 25-26, Foote does not teach the applicant's claimed automatically decomposing each image sequence into a generative model with each generative model comprising a set of model parameters that have a mean visual appearance and variance of each class in the image sequence being decomposed, by using an expectation-maximization analysis that employs a Viterbi analysis where each generative model is computed at substantially the same time that the image sequence is acquired, and an expectation step of the generalized expectation-maximization analysis maximizes a lower bound on a log-likelihood of each image frame by inferring approximations of variational parameters. Jojic also does not teach these features.

Accordingly, Foote in combination with Jojic does not teach the applicant's claim limitations. Nor does Foote in combination with Jojic recognize the advantages of the applicants' claimed invention. Namely, Foote in combination with Jojic does not teach allowing video sequences to be decomposed into a preferred number of classes in real-time. Thus, the applicants have claimed elements not taught in the cited art and which have advantages not recognized therein. Accordingly, no prima facie case of obviousness has been established in accordance with the holding of *In Re Fine*. This lack of prima facie showing of obviousness means that the rejected claims are patentable under 35 USC 103 over Foote in view of Petrovic. As such, it is respectfully requested that Claims 20-21 and 25-26 be allowed based on the previously-quoted claim language.

The applicants hereby respectfully request reconsideration of the subject application and allowance of the remaining claims at an early date.

Respectfully submitted,

Katrina A. Lyon
Reg. No. 42,821
Attorney for Applicant(s)